

Selector de Modelos

Domingo Mery

1 DEFINICIONES

• Extractor f :

Extrae m características de un fuente de datos originales \mathbf{I} , donde \mathbf{I} puede ser una imagen, una señal, etc.:

$$(\mathbf{x}, y) = f(\mathbf{I}), \quad (1)$$

en este caso \mathbf{x} es un vector de $1 \times m$ elementos, esto es las características de una muestra de la clase y (*label*). Para un conjunto de N objetos:

$$(\mathbf{X}, \mathbf{y}) = \{f(\mathbf{I}_i)\}_{i=1}^N \quad (2)$$

en este caso \mathbf{X} es una matriz de $N \times m$ elementos, cuyos *labels* se almacenan en el vector \mathbf{y} de $N \times 1$ elementos.

• Selector/Transformador de Características g :

Selecciona/Transforma las mejores p características de \mathbf{X} . La función g se aprende con datos de entrenamiento (\mathbf{X}, \mathbf{y}) y tiene hiper-parámetros θ_g (uno de ellos es p). El resultado es una función aprendida g^* :

$$g^* \leftarrow g(\mathbf{X}, \mathbf{y}, \theta_g), \quad (3)$$

de esta forma, las nuevas características es una matriz de $N \times p$ elementos, definida como:

$$\mathbf{X}^+ = g^*(\mathbf{X}). \quad (4)$$

Un ejemplo de g es SFS, en este caso los hiper-parámetros son la función de separabilidad (por ejemplo Fisher) y p . La función aprendida g^* extrae las p columnas seleccionadas de \mathbf{X} . La función g también puede ser una combinación de PCA con SFS.

• Clasificador h :

Clasifica las muestras en clases. La función h se aprende con datos de entrenamiento (\mathbf{X}, \mathbf{y}) y tiene hiper-parámetros θ_h . El resultado es una función aprendida h^* :

$$h^* \leftarrow h(\mathbf{X}, \mathbf{y}, \theta_h), \quad (5)$$

de esta manera, si queremos clasificar un nuevo conjunto de datos \mathbf{X}' con el clasificador entrenado, la predicción sería:

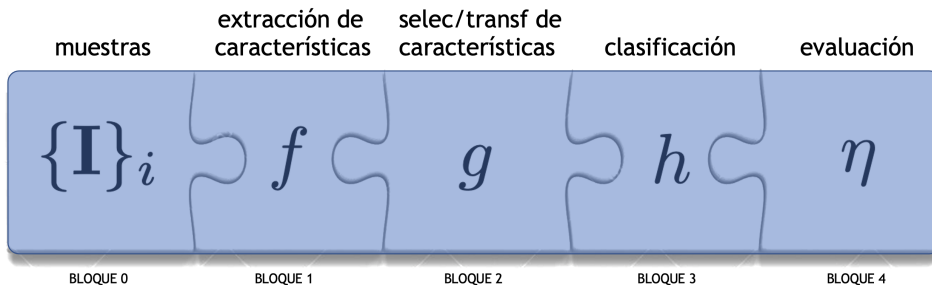
$$\hat{y} = h^*(\mathbf{X}') \quad (6)$$

Un ejemplo de h es KNN, en este caso el hiper-parámetro es el número de vecinos (por ejemplo $K = 5$).

• Separador s :

Separa las muestras. La función s separa las N muestras (\mathbf{X}, \mathbf{y}) en dos conjuntos $(\mathbf{X}_1, \mathbf{y}_1)$ y $(\mathbf{X}_2, \mathbf{y}_2)$, de N_1 y N_2 muestras respectivamente, donde $N_1 + N_2 = N$. La función s tiene como parámetros θ_s que incluye la proporción $t = N_2/N$, la definición si las filas de (\mathbf{X}, \mathbf{y}) se escogen de forma aleatoria o secuencial, y la estratificación, es decir si se mantiene o no el porcentaje de muestras por clase de los datos originales en los nuevos conjuntos de datos:

$$(\mathbf{X}_1, \mathbf{y}_1, \mathbf{X}_2, \mathbf{y}_2) = s(\mathbf{X}, \mathbf{y}, \theta_s) \quad (7)$$



2 EVALUACIÓN DE DESEMPEÑO

Dentro de las métricas de evaluación de desempeño, las más usadas son *hold-out* y *cross-val*.

• Hold-Out:

Input: muestras de entrenamiento $(\mathbf{X}_1, \mathbf{y}_1)$, muestras de prueba $(\mathbf{X}_2, \mathbf{y}_2)$, obtenidas con (7), y el clasificador (h, θ_h) .

Output: Accuracy η

Algoritmo: $\eta = \text{holdout}(\mathbf{X}_1, \mathbf{y}_1, \mathbf{X}_2, \mathbf{y}_2, h, \theta_h)$

$$h^* \leftarrow h(\mathbf{X}_1, \mathbf{y}_1, \theta_h)$$

$$\hat{y} = h^*(\mathbf{X}_2)$$

$$\eta = \text{Accuracy}(\mathbf{y}_2, \hat{y})$$

• Cross-Validation CV:

Input: muestras de entrenamiento (\mathbf{X}, \mathbf{y}) , el clasificador (h, θ_h) y el número de *folds* v .

Output: Accuracy η

Algoritmo: $\eta = \text{crossval}(\mathbf{X}, \mathbf{y}, h, \theta_h, v)$

$$\{\mathbf{X}_i, \mathbf{y}_i\}_{i=1}^v \leftarrow v \text{ folds of } (\mathbf{X}, \mathbf{y})$$

for $i = 1$ to v :

$$(\mathbf{X}_a, \mathbf{y}_a) = \{\mathbf{X}_j, \mathbf{y}_j\}_{j=1, j \neq i}^v$$

$$(\mathbf{X}_b, \mathbf{y}_b) = (\mathbf{X}_i, \mathbf{y}_i)$$

$$\eta_i = \text{holdout}(\mathbf{X}_a, \mathbf{y}_a, \mathbf{X}_b, \mathbf{y}_b, h, \theta_h)$$

$$\eta = \frac{1}{v} \sum_i \eta_i$$

3 MODELO

Un modelo se refiere a un selector/transformador de características g con parámetros θ_g y un clasificador h con parámetros θ_h . Para evaluar el desempeño de este modelo se puede usar *hold-out* o *cross-val*. En estas métricas el input es (\mathbf{X}, \mathbf{y}) obtenido de (2), y el output es el accuracy (η) .

• Evaluación del modelo usando hold-out:

$$(\mathbf{X}_1, \mathbf{y}_1, \mathbf{X}_2, \mathbf{y}_2) = s(\mathbf{X}, \mathbf{y}, \theta_s)$$

$$g^* \leftarrow g(\mathbf{X}_1, \mathbf{y}_1, \theta_g)$$

$$\mathbf{X}_1^+ = g^*(\mathbf{X}_1)$$

$$\mathbf{X}_2^+ = g^*(\mathbf{X}_2)$$

$$\eta = \text{holdout}(\mathbf{X}_1^+, \mathbf{y}_1, \mathbf{X}_2^+, \mathbf{y}_2, h, \theta_h)$$

• Evaluación del modelo usando CV:

$$g^* \leftarrow g(\mathbf{X}, \mathbf{y}, \theta_g)$$

$$\mathbf{X}^+ = g^*(\mathbf{X})$$

$$\eta = \text{crossval}(\mathbf{X}^+, \mathbf{y}, h, \theta_h, v)$$

4 SELECTOR DE MODELOS

Muchas veces se cuenta con un conjunto de n funciones selectoras/transformadoras de características: $\{g_i\}_{i=1}^n$ y un conjunto de q clasificadores: $\{h_k\}_{k=1}^q$, y se desea escoger la combinación g_i y h_k que obtenga el mejor desempeño. Para resolver este problema se pueden seguir las siguientes estrategias. En todas ellas el input es (\mathbf{X}, \mathbf{y}) obtenido de (2), y el output es el accuracy (η) y los índices (\hat{i}, \hat{k}) que indican el selector/transformador y clasificador escogidos.

• Estrategia 1: Train-1/Test-2 + hold-out

$$(\mathbf{X}_1, \mathbf{y}_1, \mathbf{X}_2, \mathbf{y}_2) = s(\mathbf{X}, \mathbf{y}, \theta_s)$$

for $i = 1$ to n :

$$g_i^* \leftarrow g_i(\mathbf{X}_1, \mathbf{y}_1, \theta_{g_i})$$

$$\mathbf{X}_1^+ = g_i^*(\mathbf{X}_1)$$

$$\mathbf{X}_2^+ = g_i^*(\mathbf{X}_2)$$

for $k = 1$ to q :

$$\eta_{ik} = \text{holdout}(\mathbf{X}_1^+, \mathbf{y}_1, \mathbf{X}_2^+, \mathbf{y}_2, h_k, \theta_{h_k})$$

$$(\hat{i}, \hat{k}) = \text{argmax}(\eta_{ik})$$

$$\eta = \eta_{\hat{i}\hat{k}}$$

• Estrategia 2: Train + cross-val

for $i = 1$ to n :

$$g_i^* \leftarrow g_i(\mathbf{X}, \mathbf{y}, \theta_{g_i})$$

$$\mathbf{X}^+ = g_i^*(\mathbf{X})$$

for $k = 1$ to q :

$$\eta_{ik} = \text{crossval}(\mathbf{X}^+, \mathbf{y}, h_k, \theta_{h_k}, v)$$

$$(\hat{i}, \hat{k}) = \text{argmax}(\eta_{ik})$$

$$\eta = \eta_{\hat{i}\hat{k}}$$

• Estrategia 3: Train-1/Val-2/Test-3 + hold-out

$$(\mathbf{X}_1, \mathbf{y}_1, \mathbf{X}_3, \mathbf{y}_3) = s(\mathbf{X}, \mathbf{y}, \theta_{s_1})$$

$$(\mathbf{X}_1, \mathbf{y}_1, \mathbf{X}_2, \mathbf{y}_2) = s(\mathbf{X}_1, \mathbf{y}_1, \theta_{s_2})$$

for $i = 1$ to n :

$$g_i^* \leftarrow g_i(\mathbf{X}_1, \mathbf{y}_1, \theta_{g_i})$$

$$\mathbf{X}_1^+ = g_i^*(\mathbf{X}_1)$$

$$\mathbf{X}_2^+ = g_i^*(\mathbf{X}_2)$$

for $k = 1$ to q :

$$\eta_{ik} = \text{holdout}(\mathbf{X}_1^+, \mathbf{y}_1, \mathbf{X}_2^+, \mathbf{y}_2, h_k, \theta_{h_k})$$

$$(\hat{i}, \hat{k}) = \text{argmax}(\eta_{ik})$$

$$(\mathbf{X}_1, \mathbf{y}_1) = \text{concatenate}(\mathbf{X}_1, \mathbf{y}_1, \mathbf{X}_2, \mathbf{y}_2)$$

$$g_i^* \leftarrow g_i(\mathbf{X}_1, \mathbf{y}_1)$$

$$\mathbf{X}_1^+ = g_i^*(\mathbf{X}_1)$$

$$\mathbf{X}_3^+ = g_i^*(\mathbf{X}_3)$$

$$\eta = \text{holdout}(\mathbf{X}_1^+, \mathbf{y}_1, \mathbf{X}_3^+, \mathbf{y}_3, h_{\hat{k}}, \theta_{h_{\hat{k}}})$$

• **Estrategia 4:** Train-1/Test-2 + cross-val

$$(\mathbf{X}_1, \mathbf{y}_1, \mathbf{X}_2, \mathbf{y}_2) = s(\mathbf{X}, \mathbf{y}, \theta_s)$$

for $i = 1$ to n :

$$g_i^* \leftarrow g_i(\mathbf{X}_1, \mathbf{y}_1, \theta_{g_i})$$

$$\mathbf{X}_1^+ = g^*(\mathbf{X}_1)$$

for $k = 1$ to q :

$$\eta_{ik} = \text{crossval}(\mathbf{X}_1^+, \mathbf{y}_1, h_k, \theta_{h_k}, v)$$

$$(\hat{i}, \hat{k}) = \text{argmax}(\eta_{ik})$$

$$\mathbf{X}_1^+ = g_{\hat{i}}^*(\mathbf{X}_1)$$

$$\mathbf{X}_2^+ = g_{\hat{i}}^*(\mathbf{X}_2)$$

$$\eta = \text{holdout}(\mathbf{X}_1^+, \mathbf{y}_1, \mathbf{X}_2^+, \mathbf{y}_2, h_{\hat{k}}, \theta_{\hat{k}})$$